

Prosiding
ANNUAL RESEARCH SEMINAR 2016
 6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

Pengaruh Penggunaan *Overlapped Character* untuk meningkatkan *Robustness* CAPTCHA

Muhammad Ezar Al Rivan

Teknik Informatika

STMIK GI MDP

Palembang, Indonesia

meedzhar@mdp.ac.id

Sehat Martinus Surya Benediktus

Teknik Informatika

STMIK GI MDP

Palembang, Indonesia

sehat.martinus@mdp.ac.id

Abstrak—CAPTCHA merupakan cara yang digunakan untuk menentukan apakah pengguna adalah manusia atau komputer. CAPTCHA berbasis teks merupakan jenis CAPTCHA yang murah, cepat dan tidak memerlukan sumber daya yang besar. OCR dapat digunakan untuk memecahkan CAPTCHA berbasis teks. Bot yang menerapkan AI dapat digunakan untuk memecahkan CAPTCHA. CAPTCHA yang diciptakan harus mudah dikenali oleh manusia namun sulit dipecahkan oleh bot. CAPTCHA yang didalamnya ada karakter yang overlapped menyulitkan bot memecahkan CAPTCHA. Penggunaan overlapped character pada CAPTCHA dapat menjadi solusi agar CAPTCHA tidak dapat dipecahkan oleh bot sehingga CAPTCHA menjadi lebih robust.

Keywords—CAPTCHA; OCR; AI; bot; overlapped character

I. LATAR BELAKANG

Completely Automated Public Turing Test to Tell Computers and Human Apart (CAPTCHA) merupakan suatu cara yang digunakan untuk membedakan manusia dan komputer. Komputer yang dimaksud yaitu adalah program komputer. CAPTCHA bekerja dengan cara memberikan tantangan yang hanya bisa dijawab oleh manusia sehingga jika pengguna benar-benar manusia maka tantangan tersebut dapat dijawab dengan benar.

Salah satu jenis CAPTCHA yaitu CAPTCHA berbasis teks. CAPTCHA ini merupakan CAPTCHA yang murah, cepat dan tidak memerlukan banyak sumber daya. Cara kerja CAPTCHA jenis ini yaitu sistem akan men-generate gambar yang terdiri dari beberapa karakter yang disusun sedemikian rupa. Pengguna diminta untuk menebak karakter yang ada di

dalam gambar tersebut.

Seiring dengan perkembangan teknologi, banyak layanan dari internet yang digunakan oleh manusia. Layanan email gratis merupakan salah satu layanan yang menggunakan CAPTCHA. Pengisian *polling online* juga disertai dengan CAPTCHA. Tujuan penggunaan CAPTCHA yaitu untuk membedakan apakah pengguna adalah manusia atau bukan. CAPTCHA dapat mencegah pemalsuan identitas yang dilakukan oleh manusia yang berbuat kriminal dengan tujuan memperoleh keuntungan dengan cara meng-*exploit* sistem. CAPTCHA juga bermanfaat untuk mencegah akses *login* secara *brute force* yaitu dengan menggunakan semua kemungkinan *password*. *Login* dengan cara *brute force* dilakukan dengan menggunakan program otomatis (*bot*). Ketika *login* gagal beberapa kali maka pengguna akan diminta untuk memasukkan CAPTCHA untuk memastikan pengguna adalah manusia bukan *bot*.

Dengan perkembangan metode pengenalan karakter dengan menggunakan *Optical Character Recognition* (OCR), maka penggunaan CAPTCHA menjadi rentan. Hal ini dikarenakan CAPTCHA yang sebelumnya hanya dapat dikenali oleh manusia saja dapat juga dikenali oleh program. OCR memiliki kemampuan mengenali setiap karakter-karakter yang muncul di dalam CAPTCHA. Penggunaan CAPTCHA menjadi tidak tepat karena memiliki kelemahan. Perkembangan *Artificial Intelligence* (AI) juga dapat digunakan untuk mengenali karakter= karakter dari sebuah gambar. Dengan adanya kemajuan AI maka CAPTCHA jenis teks memiliki kelemahan. Dengan kemajuan OCR dan AI maka penggunaan CAPTCHA menjadi lebih rentan.

Dibalik kemampuan OCR dan AI yang dapat mengenali karakter-karakter dalam CAPTCHA ternyata OCR dan AI memiliki kelemahan. Kelemahan yang dimiliki oleh OCR dan AI yaitu tidak dapat mengenali karakter-karakter yang rusak

Prosiding
ANNUAL RESEARCH SEMINAR 2016
 6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

http://ars.ilkom.unsri.ac.id

atau banyak dengan *noise*. Karakter-karakter rusak ataupun ada *noise* seringkali tidak dapat dikenali oleh OCR dan AI. Walaupun dikenali oleh hasil pengenalan bisa menjadi salah. Dengan menggunakan kelemahan yang dimiliki oleh OCR dan AI maka CAPTCHA dapat dibuat agar tidak dapat dikenali OCR dan AI. Cara-cara yang dapat dilakukan untuk merusak atau memberi *noise* diharapkan menyulitkan OCR dan AI untuk mengenali karakter-karakter dalam CAPTCHA.

Karakter-karakter rusak dan banyak *noise* juga dapat menyulitkan pengguna yang benar-benar manusia sehingga perlu mempertimbangkan aspek manusia dalam pembuatan CAPTCHA. Dalam membuat CAPTCHA diperlukan cara menyeimbangkan bagaimana cara agar CAPTCHA dapat dikenali oleh manusia namun *bot* tidak mampu untuk mengenalinya. Tingkat kesulitan memecahkan CAPTCHA harus berbanding terbalik antara manusia dan *bot*. CAPTCHA yang baik yaitu CAPTCHA yang memiliki kriteria-kriteria yaitu dapat dikenali oleh manusia dengan mudah namun sulit dikenali oleh program komputer. Tantangan yang dihadapi untuk membuat CAPTCHA yaitu bagaimana CAPTCHA yang dihasilkan dapat memenuhi kriteria dapat dikenali oleh manusia dengan mudah namun sulit dikenali oleh program komputer.

II. TINJAUAN PUSTAKA

A. Jenis CAPTCHA

CAPTCHA memiliki berbagai macam jenis. Menurut [1] jenis-jenis CAPTCHA yaitu :

- a. CAPTCHA berbasis teks
- b. CAPTCHA berbasis gambar
- c. CAPTCHA berbasis audio
- d. CAPTCHA berbasis video
- e. CAPTCHA berbasis puzzle

Masing-masing jenis CAPTCHA memiliki kelemahan. Kelemahan jenis CAPTCHA menurut [1] yaitu

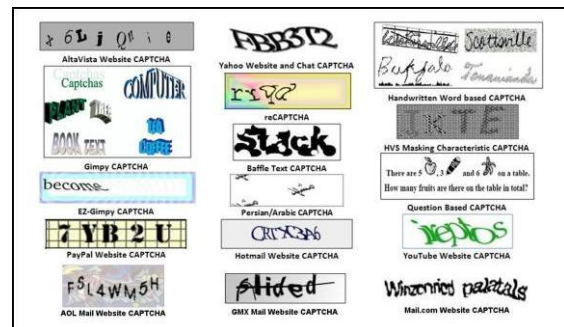
No.	Jenis CAPTCHA	Kelemahan
1.	Teks	Kesulitan pengguna mengidentifikasi karakter, dapat dikenali oleh OCR
2.	Gambar	Kesulitan pengguna yang memiliki kemampuan melihat yang rendah mengidentifikasi gambar
3.	Audio	Kesulitan bagi yang tidak mengerti bahasa Inggris dan

		adanya karakter yang memiliki bunyi sama
4.	Video	Ukuran video yang besar sehingga bermasalah saat di-download
5.	Puzzle	Memerlukan waktu yang tidak singkat

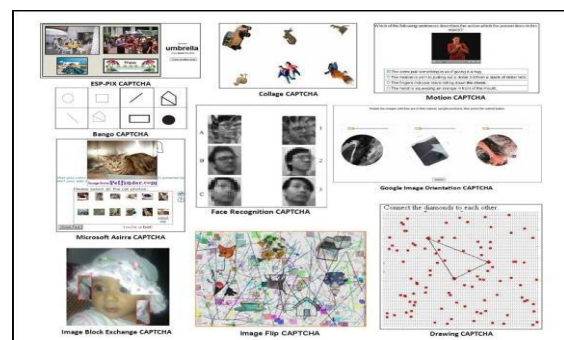
B. Penggunaan CAPTCHA

Menurut [2] CAPTCHA dapat digunakan pada aplikasi-aplikasi seperti :

- a. Mengisi *form* pada *polling online* untuk mencegah *bot* mengisi poling
- b. Layanan email gratis
- c. Mencegah halaman web terindeks oleh *search engine bots*
- d. Pencegahan terhadap *worm* dan *spam*



Gambar 1. Contoh CAPTCHA berbasis teks [3]



Gambar 2. Contoh CAPTCHA berbasis gambar [3]

Prosiding
ANNUAL RESEARCH SEMINAR 2016
6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

III. PENELITIAN TERKAIT

Penelitian yang dilakukan oleh [2] melakukan pemecahan CAPTCHA dengan 4 tahapan yaitu *preprocessing*, *segmentation*, *feature extraction* dan *character recognition*. *Preprocessing* dilakukan untuk memperbaiki gambar dengan teknik-teknik dalam pengolahan citra seperti membuang *background*, membuang *noise*, dan lain sebagainya agar hasil *segmentation* menjadi lebih baik.

Langkah selanjutnya yaitu dengan melakukan *segmentation*. *Segmentation* dilakukan agar karakter-karakter dalam gambar dapat diambil dan menjadi karakter tunggal. Karakter ini kemudian diambil fiturnya dengan cara *feature extraction*. Lalu masing-masing fitur akan dikenali melalui proses *character recognition*. Dari penelitian ini dapat diketahui bahwa tahapan-tahapan yang dilakukan oleh [2], CAPTCHA dapat dipecahkan dengan sukses. Metode yang diusulkan lebih baik dibandingkan OCR karena OCR tidak mampu mengenali karakter dengan *background* yang mengalami distorsi.

Penelitian lain yang dilakukan oleh [4]. Pemecahan CAPTCHA dilakukan dengan menggunakan teknik-teknik dalam pengolahan citra. Hasil yang diperoleh dari penelitian tersebut, hampir 100% CAPTCHA dapat dipecahkan oleh metode yang diusulkan.

Penelitian terkait lainnya seperti yang dilakukan oleh [5]. Pada penelitian yang dilakukan metode yang diusulkan yaitu *reverse engineering*. Proses dilakukan dengan membalikan proses dalam penciptaan CAPTCHA. CAPTCHA diciptakan dengan melalui beberapa proses. Dengan mengetahui proses-proses penciptaan CAPTCHA maka dapat diketahui komponen-komponen yang ditambahkan dalam setiap proses penciptaan CAPTCHA. Proses *reverse engineering* dapat dilakukan dengan *layering* yaitu mengetahui pola-pola yang terbentuk dari gambar. Proses berikutnya yaitu dengan mengetahui *background* dari gambar. Proses berikutnya mengetahui jenis dari *noise* yang ada pada gambar. Proses berikutnya dengan mengetahui jenis teks yang digunakan pada gambar CAPTCHA. Proses terakhir yaitu dengan mengetahui proses transformasi pada setiap karakter di gambar. Dari hasil penelitian tersebut diketahui bahwa metode yang digunakan oleh [5] dapat memecahkan CAPTCHA dengan sukses.

IV. METODE YANG DIUSULKAN

Berdasarkan hasil penelitian yang dilakukan, [5] memberikan rekomendasi dalam membuat CAPTCHA berbasis teks. Adapun beberapa rekomendasi yaitu :

- a. Dalam membuat CAPTCHA sebaiknya menggunakan Transformasi Non Linear
- b. CAPTCHA menggunakan karakter yang berwarna
- c. CAPTCHA menggunakan jenis karakter yang lebih banyak seperti *lowercase*, *uppercase*, angka dan tanda baca
- d. Pengisian jawaban CAPTCHA harus dibatasi
- e. Teks yang digunakan dalam gambar dibuat mirip dengan *background*-nya
- f. CAPTCHA menggunakan karakter yang bentuknya terpotong
- g. CAPTCHA menggunakan *overlapped character*

Berdasarkan rekomendasi tersebut, eksperimen yang akan dilakukan dalam pembuatan CAPTCHA di penelitian ini yaitu sebagai berikut.

- a. Membuat CAPTCHA yang terdiri dari karakter *lowercase* dan *uppercase*.
- b. Karakter dalam CAPTCHA yang digunakan yaitu angka dan huruf.
- c. Banyaknya karakter dalam CAPTCHA yaitu 6.
- d. CAPTCHA yang diciptakan akan disusun sedemikian rupa agar ada *overlapped character*.
- e. *Background* dari gambar akan dibuat *noise*.
- f. Masing-masing karakter akan diberi warna berbeda-beda dengan cara diacak.
- g. Posisi masing-masing karakter akan diacak.

CAPTCHA yang dibuat nantinya akan diujikan. CAPTCHA akan diujikan kepada manusia dan *bot*. Untuk pengujian terhadap manusia dilakukan dengan menggunakan sistem. Manusia diminta untuk menjawab CAPTCHA yang dihasilkan oleh sistem. CAPTCHA yang dihasilkan oleh sistem akan disimpan dalam bentuk gambar. Selain itu jawaban yang benar dari CAPTCHA juga akan disimpan dalam *database*. Di dalam pengujian yang dilakukan terhadap manusia, jawaban manusia juga akan disimpan dalam *database* beserta dengan status jawaban apakah benar atau salah. Jawaban yang benar dan salah dihitung secara keseluruhan dan per karakter.

Pengujian yang dilakukan terhadap bot dilakukan dengan memasukkan gambar yang disimpan hasil dari *generate* sistem pada pengujian yang dilakukan terhadap manusia. Gambar-gambar CAPTCHA akan di *load* ke *bot*. *Bot* kemudian akan memecahkan CAPTCHA yang sama dengan CAPTCHA yang telah diujikan kepada manusia. *Bot* akan memberikan jawaban hasil pemecahan CAPTCHA. Hasil pemecahan CAPTCHA ini akan dibandingkan dengan jawaban yang sebenarnya yang

Prosiding
ANNUAL RESEARCH SEMINAR 2016
6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

telah disimpan pada saat CAPTCHA di *generate*. Proses evaluasi dilakukan dengan cara yang sama dengan pengujian yang dilakukan terhadap manusia. Jawaban dari *bot* akan dibandingkan dengan jawaban sebenarnya. Status benar dan salah dari jawaban *bot* dihitung secara keseluruhan dan per karakter yang ada di CAPTCHA.

V. HASIL DAN PEMBAHASAN

Adapun hasil dari eksperimen ini dapat dilihat pada tabel berikut.

TABEL II. Perbandingan Hasil Pengujian Captcha Antara Manusia Dan Bot

	Manusia	Bot
Benar	143	0
Salah	17	160
Jumlah	160	160

TABEL III. Perbandingan Per Karakter Terhadap Manusia

	Karakter ke-					
	1	2	3	4	5	6
Benar	160	156	160	158	152	158
Salah	0	4	0	2	8	2
Jumlah	160	160	160	160	160	160

TABEL IV. Perbandingan Per Karakter Terhadap Bot

	Karakter ke-					
	1	2	3	4	5	6
Benar	160	160	160	160	17	129
Salah	0	0	0	0	143	31

Jumlah	160	160	160	160	160	160

Secara keseluruhan, CAPTCHA yang dihasilkan tidak dapat dipecahkan oleh *bot*. Tidak ada satupun CAPTCHA yang dihasilkan yang dipecahkan *bot* dengan benar. Namun, jika dilihat dari jawaban manusia, terdapat 10,625% CAPTCHA yang tidak berhasil dipecahkan oleh manusia.

Jika dihitung per karakter, *bot* memiliki kemampuan memecahkan CAPTCHA untuk 4 karakter pertama dengan benar. Jika dilihat pada karakter ke-5 dan ke-6, *bot* tidak bisa memecahkan CAPTCHA dengan benar. Untuk karakter ke-5 hanya 10,625% karakter pada CAPTCHA yang dapat dipecahkan oleh *bot*. Untuk karakter ke-6 ada 80,625% karakter pada CAPTCHA yang dapat dipecahkan oleh *bot*. Penyebab kesalahan menjawab karakter ke-5 dan ke-6 ini dikarenakan adanya karakter yang saling *overlapped*. Karakter *overlapped* ini menyulitkan *bot* memecahkan CAPTCHA.

Untuk hasil evaluasi terhadap manusia, kesalahan untuk setiap karakter lebih bervariasi. Kesalahan terbesar terletak pada karakter ke-5 yaitu sebanyak 5 % kemudian karakter ke-2 sebanyak 2,5% dan karakter ke-4 dan karakter ke-6 masing-masing 1,25%. Kesalahan manusia menjawab pada karakter ke-5 dan ke-6 disebabkan oleh karakter yang *overlapped* yang tidak dapat dilihat dengan mudah oleh manusia. Untuk karakter ke-2 dan ke-4 lebih disebabkan karena pewarnaan dari karakter yang sulit dibedakan manusia sehingga terjadi kesalahan dalam menjawab.

Berdasarkan hasil eksperimen, secara umum model CAPTCHA yang diusulkan tidak dapat dipecahkan oleh *bot*. Hal ini disebabkan adanya faktor yang mempersulit *bot* memecahkan CAPTCHA yaitu adanya karakter *overlapped*. Namun demikian ternyata ada sebagian kecil dari manusia yang juga tidak dapat menjawab CAPTCHA dengan benar yang diakibatkan oleh karakter yang *overlapped*. Dapat disimpulkan bahwa karakter *overlapped* bisa digunakan pada CAPTCHA agar *bot* tidak dapat memecahkan CAPTCHA namun perlu disusun agar karakter *overlapped* ini dapat dengan mudah dikenali oleh manusia. Selain itu, faktor pewarnaan karakter menjadi faktor yang mempersulit manusia menjawab CAPTCHA karena keterbatasan penglihatan manusia terhadap warna-warna tertentu. Pewarnaan karakter bisa disesuaikan agar dapat dikenali oleh manusia.